Robust Estimation Using Influence Function As Error

by
Hai-Li Wang & David Hinkley
University of Minnesota
Technical Report No. 337
April, 1979

Research supported in part by NSF Grant MCS-77-00959

Robust Estimation Using Influence Function As Error

Hai-Li Wang & David Hinkley

University of Minnesota

Summary

We consider robust estimates using influence function values as generalized errors, where the influence function is that of an initial standard estimate. The general approach is illustrated in the case of estimating a correlation coefficient.

Key words: Robust estimation; influence function; correlation.

1. INTRODUCTION

The development of robust estimates for location parameters and linear model parameters has been based on the recognition that least-squares estimates are particularly sensitive to non-normal errors, or abnormally large errors, in the relevant observation model. The distribution-free approach espoused by Bickel (1976), Huber (1972, 1973) and Hampel (1974), among others, makes much use of the influence function as a means of assessing the sensitivity and efficiency of various classes of estimates for linear parameters. The sensitivity of least-squares methods is evidenced by unboundedness of the influence function.

Robust estimation theory for the linear model is made relatively easy by the fact that one typically assumes additive error models, with symmetrically distributed errors. If the model for observations y is

$$y = m(\theta,u) + e \tag{1.1}$$

where u is a covariate, then a natural class of criteria for estimation based on data $(u_1,y_1),\ldots,(u_n,y_n)$ is

$$Q(\theta) = \Sigma\rho(y_i - m(\theta,u_i)) \quad , \tag{1.2}$$

where for least-squares $\rho(e) = e^2$. Robust estimates derived by minimizing $Q(\theta)$ are referred to as M-estimates, and have the characteristic of bounded sensitivity (Hampel, 1974) if the derivative of $\rho(e)$ is bounded.

In the present paper we show how the M-estimation methods may be applied to more general problems than (1.1). The methods may be thought of as improving on tentative estimators by construction of approximate linear models. The general idea is explained simply as follows. Suppose that T_n is a regular estimate of the simple parameter θ , where T_n is chosen to match certain ideal assumptions for the data, but so that T_n is meaningful under wider conditions. Then usually a leading-term asymptotic expansion of T_n will look like

$$T_n = \theta + n^{-1}\Sigma a_\theta(X_j)$$

Thus T_n is approximately an average of quantities $Y = \theta + a_\theta(X)$, where moreover $Ea_\theta(X) = 0$. If the Y_j were observable, we might then replace T_n by an estimate minimizing (1.2) for suitable $\rho(.)$ with $m(\theta,u_i) = \theta$. There are two ways to overcome the fact that Y is unobservable, the first of which is self-evident because

$$Y - m(\theta,u) = a_\theta(X) \quad . \tag{1.3}$$

The second way is to estimate Y by the "pseudo-value"

$$\hat{Y} = T_n + \hat{a}_\theta(X) \; , \tag{1.4}$$

where some version of the jackknife is used to compute $\hat{a}_\theta(X)$. This second possibility, described briefly by Hinkley (1977,1978), will be discussed elsewhere. The present paper deals with the first approach where (1.3) is used in conjunction with the criterion (1.2).

In Section 2 we outline the theory of M-estimates using $a_\theta(X)$ as error term when θ is the only parameter. Section 3 describes the corresponding theory when nuisance parameters are present. The simple case of correlation estimation is used in Section 4 to illustrate various features of the methods, including their small-sample distributional properties. Section 5 contains a brief summary.

2. THEORY FOR THE SINGLE PARAMETER CASE

We shall confine our attention to the case where T_n is a differentiable function $t(\tilde{F}_n)$ of the sample distribution function

$$\tilde{F}_n(x) = n^{-1} \sum_{j=1}^{n} I(x-X_j) \quad ,$$

where $X_1,\ldots,X_n$ are independent random variables each with distribution function $F = F_\theta$ depending on the single parameter θ . Slightly more generality could be obtained by allowing T_n to depend on an auxiliary real variable z_n so that $T_n = t(\tilde{F}_n, z_n)$ and our subsequent discussion could be generalized accordingly. By taking an appropriate definition of von Mises differentiability of $t(F)$, we have the one-term Taylor expansion

$$T_n = t(\tilde{F}_n) = t(F) + \int t_1(F;y)\{d\tilde{F}_n(f) - dF(y)\} + o(\| \tilde{F}_n-F \|) \quad ,$$

or more simply

$$T_n = \theta + n^{-1} \Sigma a_\theta(Y_j) + o_p(n^{-\frac{1}{2}}) \quad .$$

Here

$$a_\theta(y) = t_1(F;y) - \dot{E}\{ t_1(F;y)\}$$

is the centered von Mises derivative of T_n, more usually called the influence function (Hampel, 1974; unpublished Harvard thesis by J. Reeds).

Our aim is to improve upon T_n by filtering its influence function through a bounded criterion, specifically by forming the estimate S_n to minimize

$$\sum_{j=1}^{n} \rho\{ a_s(X_j)\} \quad .$$

If $\psi(u) = d\rho(u)/du$, then we let S_n be a solution to

$$n^{-1}\sum_{j=1}^{n} \psi\{ a_s(X_j)\} = \int\psi\{ a_s(x)\}d\tilde{F}_n(x) = 0 \quad . \tag{2.1}$$

Under certain conditions the solution will be unique. We shall suppose that $\psi(.)$ is a continuous odd function, $\psi(-u) = -\psi(u)$. Then if $s=\theta^*$ is the solution to

$$\int\psi\{ a_s(x)\}dF_\theta(x) = 0 \quad , \tag{2.2}$$

S_n will converge to θ^* almost surely because of the strong convergence of $\tilde{F}_n$ to F_θ. A sufficient condition for $\theta^* = \theta$, that is consistency of S_n, is that $a_\theta(X)$ have a symmetric distribution about its zero mean value. Proofs of these statements are complete parallels of those for likelihood equations and will not be given here.

The behaviour of the estimate S_n in large samples is most easily understood by working with equation (2.2), which gives an implicit definition of the function $s(F)$ such that $S_n = s(\tilde{F}_n)$. In particular, the influence function $s_1(F;x)$ may be derived as follows. Replace F by $F + \varepsilon(G-F)$ in (2.2), write S explicitly as $s(F + \varepsilon(G-F))$, and then differentiate with respect to ε at $\varepsilon=0$. The result is, with $s = s(F)$ and $\dot{\psi}(u) = d\psi(u)/du$,

$$\int\psi\{ a_s(x)\}\{ dG(x)-dF(x)\} + \iint s_1(F;x)\frac{\partial}{\partial s} a_s(y)\dot{\psi}\{a_s(y)\}dF(y)\{ dG(x)-dF(x)\} = 0 ,$$

which implies that in general

$$s_1(F;x) = \frac{\psi\{a_\theta*(x)\}}{-\int\dot{\psi}\{a_\theta*(y)\}\frac{\partial}{\partial \theta*}a_{\theta*}(y)\ dF_\theta(y)} . \tag{2.3}$$

One important feature of this result is its implication that a bounded ψ function gives an estimate S_n with bounded influence function.

The limiting distribution properties of S_n may be obtained by formal expansion of the estimating equation (2.1) in Taylor series, or by applying a standard delta theory for von Mises differentiable statistical functions, either of which lead to the result that

$$\sqrt{n}(S_n - \theta^*)$$

has a limiting distribution with variance $\tau^2 = \text{var}\{s_1(F;X)\}$ Primary references for proof of this result are the unpublished theses by the first author and by J. Reeds. It is of course an important requirement that the denominator of the expression (2.3) be non-zero, since otherwise a second-order asymptotic theory would be required.

There are various ways to estimate the variance τ^2 in the approximating normal distribution of S_n , all of them being of the form

$$\hat{\tau}^2 = n^{-1}\sum_{j=1}^{n}\{\hat{s}_1(F;X_j)\}^2 .$$

Perhaps the simplest approach is to use the infinitesimal jackknife estimate

$$\hat{s}_1(F;x) = s_1(\tilde{F}_n;x) = \frac{\psi\{a_{S_n}(x)\}}{-\frac{1}{n}\sum_{j=1}^{n}\dot{\psi}\{a_{S_n}(X_j)\}\frac{\partial}{\partial S_n}a_{S_n}(X_j)}$$

Other methods may be based on the ordinary jackknife. The practical advantage of the infinitesimal jackknife is that it requires only one evaluation of T and S. It may be verified that

$$\hat{\tau}^2 = \int\{ s_1(\tilde{F}_n;x)\}^2 d\tilde{F}_n(x) \to \tau^2 = \int\{s_1(F;x)\}^2 dF(x) \ , \tag{2.4}$$

almost surely, and thence we may conclude that

$$\sqrt{n}(S_n - \theta^*)/\hat{\tau}$$

is approximately standard normal for large n.

Note that in estimating s_1 we replaced θ(or θ^*) by S_n, not T_n. This involves estimation of $a_\theta(x) = t_1(x)$ with $\theta = S_n$, which will give a more robust measure of the sensitivity of T_n to x than $a_{T_n}(X)$.

3. THEORY FOR THE NUISANCE PARAMETER CASE

It is usually the case that the influence function of T_n will depend both on θ and on a nuisance parameter w, say, which is not of direct interest, but nevertheless is unknown. Thus, in the correlation case to be discussed later w contains means and variances. For simplicity of exposition we shall assume θ and w both one dimensional, and denote the influence function of T_n by $a_{\theta,w}(x)$.

The simplest way to generalize estimating equation (2.1) is to substitute an estimate $w_n = w(\tilde{F}_n)$, so that S_n is a solution to

$$\int\psi\{a_{s,w(\tilde{F}_n)}(x)\}d\tilde{F}_n(x) = 0 \ . \tag{3.1}$$

The corresponding definition of $s(F)$ is the solution to (3.1) with F in a place of $\tilde{F}_n$, and then $S_n \to \theta^* = s(F)$ as $n\to\infty$. We shall assume that a unique solution $s(F)$ exists. One alternative to (3.1) is to solve the equation

$$\int \psi\{a_{s,w(\tilde{F}_n)}(x)\} d\tilde{F}_n(x) = 0 \tag{3.2}$$

jointly with another equation set up for estimating w, but we shall not discuss this approach explicitly.

It is evident from the definition (3.1) that if ψ is differentiable and if $w(F)$ is differentiable, then S_n may be written as $s(\tilde{F}_n)$ such that s is differentiable and $S_n \to \theta^*(w)$ as $n\to\infty$. If $a_{\theta,w}(X)$ has a symmetric distribution and if $w(F_n) \to w$, then $\theta^*(w) = \theta$ and S_n is consistent. We shall assume that this is so. However, even with this symmetry assumption the influence function of S_n will not generally be the same as (2.3). In fact we find that

$$s_1(F;x) = \frac{\psi\{a_{\theta,w}(x)\} + c_{\theta,w}\, w_1(F;x)}{d_{\theta,w}} \tag{3.3}$$

where $w_1(F;x)$ is the influence function for $\hat{w}_n$ and

$$c_{\theta,w} = E[\frac{\partial a_{\theta,w}(X)}{\partial w} \dot{\psi}\{a_{\theta,w}(X)\}] \tag{3.4}$$

$$d_{\theta,w} = E[-\frac{\partial a_{\theta,w}(X)}{\partial \theta} \dot{\psi}\{a_{\theta,w}(X)\}] \quad . \tag{3.5}$$

In general s_1 is now unbounded if w_1 is unbounded.

Alternative forms for $c_{\theta,w}$ and $d_{\theta,w}$ are found by differentiating $E[\psi\{a_{\theta,w}(X)\}] = 0$ with respect to θ and w. If $f_{\theta,w}$ is the density of $F_{\theta,w}$ and if we write

$$\dot{\ell}_\theta(x) = \frac{\partial \log f_{\theta,w}(x)}{\partial \theta} \quad , \quad \dot{\ell}_w(x) = \frac{\partial \log f_{\theta,w}(x)}{\partial w} \quad ,$$

then

$$c_{\theta,w} = -\mathrm{cov}[\psi\{a_{\theta,w}(X)\} , \dot{\ell}_w(X)]$$

$$d_{\theta,w} = \mathrm{cov}[\psi\{a_{\theta,w}(X)\} , \dot{\ell}_\theta(X)] \quad .$$

As in Section 2, standard asymptotic theory shows that $\sqrt{n}(S_n-\theta)$ converges to a $N(0,\tau^2)$ variable with $\tau^2 = \mathrm{Var}\{s_1(F;X)\}$. This variance can be estimated by the sample variance of a consistent estimate of s_1 , as before.

We have noted that s_1 will be unbounded if w_1 is unbounded, unless of course $c_{\theta,w}$ is zero. Although it would typically be true that $|s_1| << |t_1|$ as $|t_1|$ increases, it nevertheless makes sense to consider robust estimates for w . Suppose, for example, that w is a scale parameter. Now a robust estimate of scale, such as median absolute deviation, will typically not converge to w , but rather to $w^* \neq w$, say. In this situation consistency of S_n , the solution to (3.1), will require that for $w^* = \mathrm{plim}\ w(\tilde{F}_n)$

$$\int \psi\{a_{\theta,w^*}(x)\} dF_\theta(x) = 0 \quad . \tag{3.6}$$

Curiously this turns out to be satisfied in a wide variety of cases, including the correlation case discussed in Section 4.

Thus the preceding asymptotic theory, which assumed symmetry of $a_{\theta,w}(X)$ and consistency of $\hat{w}_n = w(\tilde{F}_n)$, does hold more generally. This permits the use of robust estimates $\hat{w}_n$ which will result in bounded sensitivity of S_n .

4. EXAMPLE: THE CORRELATION COEFFICIENT

A simple non-trivial illustration of the above theory is provided by the case of the bivariate correlation coefficient. Thus X is bivariate (Y,Z), and for correlation coefficent ρ we take $\theta = \tanh^{-1}\rho$. As basic estimate we take

$$T_n = \tanh^{-1} R_n \quad ,$$

with R_n the normal-theory product-moment estimate. A fairly straightforward calculation (Devlin _et al_, 1975) gives

$$a_{\theta,w}(y,z) = \{\tilde{y}\tilde{z} - \tfrac{1}{2}\rho(\tilde{y}^2 + \tilde{z}^2)\}/(1-\rho^2) \tag{4.1}$$

with $\tilde{y} = (y-\mu_Y)/\sigma_Y$ and $\tilde{z} = (z-\mu_Z)/\sigma_Z$ the standardized values of y and z. Here $w = (\mu_Y, \mu_Z, \sigma_Y, \sigma_Z)$.

If w is known the theory of Section 2 applies, and S_n will be consistent if X has an elliptically symmetric distribution with density of the form

$$f(x) = k(|\Sigma|)\ \phi\{(x-\mu)^T\Sigma^{-1}(x-\mu)\} \quad , \tag{4.2}$$

where

$$\Sigma = \begin{pmatrix} \sigma_Y^2 & \rho\sigma_Y\sigma_Z \\ \rho\sigma_Y\sigma_Z & \sigma_Z^2 \end{pmatrix} .$$

This, together with some other results to be described shortly, is proved in the Appendix.

For bivariate normal data, $a_{\theta,w}(X)$ has constant variance 1, and is distributed as the product of two independent standard normal variables. These facts can be used as a basis for choosing a suitable ψ function.

For illustration we choose

$$\psi(a) = \begin{cases} b\ \mathrm{sgn}(a) & |a| \geq b \\ a & |a| < b \end{cases} , \tag{4.3}$$

with $1 \leq b \leq 2$. In the normal case the probabilities of $|a| \geq b$ are close to 0.8, 0.9 and 0.95 for $b = 1.0$, 1.5 and 2.0 respectively.

When w is unknown we have the complication of estimating four nuisance parameters, discussed in Section 3. However, if the distribution is of the form (4.2), then the constant vector $c_{\theta,w}$ in (3.3) is zero, so that the influence function s_1 does not depend on w_1 at all; see Appendix. Also, for mixtures of normal densities, the constant $d_{\theta,w}$ in (3.3) is independent of θ and w . If s_1 no longer involves w_1 , then S_n will not be influenced by choice of estimate for w in infinitely large samples. In small samples the choice of $\hat{w}_n$ could still be important.

The normal-theory estimates of w will converge to w , so that S_n will be consistent using such estimates. For other estimates of w we need to verify that (3.6) is satisfied; we are now assuming (4.2). Most robust estimates of μ_Y and μ_Z are unbiased, so that the first two components of w^* and w agree. If the same robust method is used for each scale parameter, then their estimates will converge to $\gamma\sigma_Y$ and $\gamma\sigma_Z$ for some γ . But then, by (4.1),

$$a_{\theta,w^*}(X) = a_{\theta,w}(X)/\gamma^2 , \tag{4.4}$$

so that (3.6) will be satisfied and S_n will be consistent.

Our numerical results will be based on normal distributions and mixtures thereof, so that the influence function (3.3) of S_n is ,

$$s_1(x) = \psi\{a_{\theta,w}(x)\}/d \qquad (4.5)$$

where $\underline{a}$ is given by (4.1), ψ by (4.3), and d defined by (3.5) is constant. For the normal case Monte Carlo simulation gives $d = 0.52$, 0.68 and 0.79 at $b = 1.0$, 1.5, and 2.0. Thus $|s_1|$ is bounded by approximately 2.0, 2.25 and 2.5 for $b = 1$, 1.5 and 2 respectively. These bounds are rough guides in general, since d does depend on F.

As we have said, the small-sample properties of S_n may well depend on the choice of estimate for w. We consider two estimates. First, $\bar{w}_n$ is the usual non-robust normal-theory maximum likelihood estimate. The other estimate, denoted $\dot{w}_n$, is the usual robust median-type estimate defined by

$$\dot{y} = \text{median } y,\ \dot{z} = \text{median } z,\ \dot{s}_y = \frac{\text{median}|y-\dot{y}|}{0.6745},\ \dot{s}_Z = \frac{\text{median } |z-\dot{z}|}{0.6745} .$$

Although $\dot{s}_y$ and $\dot{s}_Z$ are consistent only for normal data, (4.4) holds so that S_n will be consistent.

The bounded influence function for $\dot{w}_n$, while simple, does involve the density f at the quartiles, which is not simple to estimate. The unbounded influence function for $\bar{w}_n$ is

$$\bar{w}_1(x) = (\sigma_Y\tilde{y},\ \sigma_Z\tilde{z},\ \tfrac{1}{2}\sigma_Y(\tilde{y}^2-1)\ ,\ \tfrac{1}{2}\sigma_Z(\tilde{z}^2-1))\ ,$$

which is easy to estimate. This is relevant because we wish to estimate the standard error of S_n via estimates of $s_1(x)$, and we might estimate the general form (3.3) rather than the special form (4.5) out of ignorance of the latter. We shall look at both in the following discussion. For example, if $\bar{w}_n$ is used in the calculation of S_n we shall estimate (4.5) by

$$\hat{s}_1(x) = \psi\{a_{s_n,\bar{w}_n}(x)\} \Big/ \left[-n^{-1} \sum_{j=1}^{n} \left. \frac{\partial a_{\theta,\bar{w}_n}(X_j)}{\partial \theta} \right|_{\theta=s_n} \psi\{a_{s_n,\bar{w}_n}(X_j)\} \right] ,$$

where of course ψ , the derivative of (4.3), is one or zero. The variance of S_n will be estimated by

$$\frac{1}{n^2} \sum_{j=1}^{n} \{\hat{s}_1(X_j)\}^2 \quad .$$

When (3.3) is used, corresponding estimates of $\bar{w}_1(x)$ and $c_{\theta,w}$ are needed.

A small-scale Monte Carlo experiment has been run to examine the small-sample behavior of S_n and its standard error estimates for normal and contaminated normal distributions. In all cases a two-step Newton-Raphson iteration from $s=T_n$ is used to solve (3.1) for S_n . In the experiment we compare the uses of $\bar{w}_n$, $\dot{w}_n$ and true w in obtaining S_n , and the corresponding estimates of $var(S_n)$ based on (4.5). We also include the estimate of $var(S_n)$ based on (3.3) when $\bar{w}_n$ is used. Table 1 contains a summary of the results for $n=20$, $\rho=0.5$ based on 500 samples from mixtures of $N(\mu,\Sigma)$ and $N(\mu,9\Sigma)$, the latter having proportions 0, 0.1 and 0.2. Truncation values $b = 1.0$, 1.5 and 2.0 are used for (4.3).

The first part of the table contrasts variations of S_n and T_n , The second part of the table concerns coverage frequencies of confidence intervals $S_n \pm 2\hat{\tau}/\sqrt{n}$, which should be approximately 95% if the large-sample theory is reliable.

Table 1. Monte Carlo results for estimates T_n and S_n of $\theta = \tanh^{-1}\rho$. Sampled distributions are bivariate normal with np values multipled by 3. S_n is defined by equations (3.1) and 4.3) with b=1, 1.5, 2. Case: n=20, ρ=0.5, 500 samples

(a) Variances of estimates

		contaminating proportion p					
		0		0.1		0.2	
	estimate	T_n	S_n	T_n	S_n	T_n	S_n
nuisance parameter value	true w	0.059	(*).064,.057,.055	.140	.066,.063,.060	.139	.082,.074,.072
	$\bar{w}_n$		.076,.068,.064		.074,.077,.083		.098,.096,.104
	$\dot{w}_n$		.065,.061,.060		.067,.063,.063		.099,.083,.086

(b) Percent coverage by nominal 95% confidence intervals for θ

	contaminating proportion p		
estimate; method for s_1	0	0.1	0.2
T_n	91.2	84.0	82.6
S_n with w; egn.(4.5)	(*)92.4,93.4,92.4	91.6,92.8,92.0	86.0,89.0,88.4
S_n with $\bar{w}_n$; egn.(4.5)	92.2,91.0,92.4	94.8,92.4,90.4	88.0,89.0,88.0
" ; egn.(3.3)	94.4,93.2,93.4	95.2,94.8,93.6	94.0,93.8,92.8
S_n with $\dot{w}_n$; egn.(4.5)	91.8,92.4,92.6	90.8,89.6,91.2	84.0,86.6,86.4

(*) values correspond to b=1, 1.5, 2 in that order

The results are quite encouraging. From the point of view of precision, there is a slight general preference for using the median-type estimate $\acute{w}_n$ for location and scale unknowns. Confidence limits based on the normal approximation seem to be quite reliable, particularly when the general form of infinitesimal jackknife is used. Similar results are obtained for other values of ρ at $n=20$.

5. CONCLUDING REMARKS

The estimation method we have described is a general method suitable for robustifying conventional estimates. The generality is an important property, in that for some problems there will not be obvious classes of robust estimates. The method has natural extensions to non-homogeneous problems, including multiple regression as suggested by Hinkley (1977) and discussed in the first author's Ph.D. thesis. In particular cases, such as the correlation problem in Section 4, problem-specific robust estimates are available and should be considered, of course.

A similar approach to the one we have discussed here would be based on order statistics of the $a_{\theta,w}(X_j)$. For example one might consider adjustments to T_n using trimmed means of the estimates $a_{T_n,\hat{w}_n}(X_j)$. This is essentially the same as using trimmed means of jackknife pseudo-values, which are examined by Hinkley & Wang (1979), and leads to estimates with unbounded influence functions when T_n is not a linear statistic.

Much of the work described in this paper is based on material in the first author's Ph.D. thesis at the University of Minnesota.

ACKNOWLEDGEMENT

Work on this paper was supported in part by the National Science Foundation.

REFERENCES

Bickel, P.J. (1976) Another look at robustness: a review of reviews and some new developments. Scand. J. Statist., 3, 145-157.

Devlin, S.J., Gnanadesikan, R. & Kettenring, J.R. (1975) Robust estimation and outlier detection with correlation coefficients. Biometrika, 62, 531-545.

Hampel, F. (1974) The influence curve and its role in robust estimation. J. Amer. Statist. Ass., 69, 383-393.

Hinkley, D.V. (1977) On jackknifing in unbalanced situations. Technometrics, 19, 285-291.

Hinkley, D.V. (1978) Improving the jackknife with special reference to correlation estimation. Biometrika, 65, 13-21.

Hinkley, D.V. & Wang, H. (1979) A trimmed jackknife. (Submitted for publication to J.R.S.S.B.)

Huber, P.J. (1972) Robust statistics: a review. Ann. Math. Statis, 43, 1041-1067.

Huber, P.J. (1973) Robust regression: asymptotics, conjectures and Monte Carlo. Ann. Stat., 1, 799-821.

APPENDIX

We give here some details of the theory for the correlation example in Section 4. With parameter $\theta = \tanh^{-1}\rho$, the influence function of the normal-theory estimate at $x = (y,z)$ is

$$a_{\theta,w}(x) = \{\tilde{y}\tilde{z} - \frac{1}{2}\rho(\tilde{y}^2+\tilde{z}^2)\}/(1-\rho^2) ,$$

where $\tilde{y} = (y-\mu_Y)/\sigma_Y$, $\tilde{z} = (z-\mu_Z)/\sigma_Z$ and $w = (\mu_Y,\mu_Z,\sigma_Y,\sigma_Z)$. The estimates S discussed in the paper have general influence function (3.3), which here involves the partial derivatives

$$\frac{\partial a}{\partial \mu_Y} = -\frac{(\tilde{z}-\rho\tilde{y})}{\sigma_Y(1-\rho^2)} , \quad \frac{\partial a}{\partial \mu_Z} = -\frac{(\tilde{y}-\rho\tilde{z})}{\sigma_Z(1-\rho^2)}$$

$$\frac{\partial a}{\partial \sigma_Y} = +\tilde{y}\frac{\partial a}{\partial \mu_Y} , \quad \frac{\partial a}{\partial \sigma_Z} = +\tilde{z}\frac{\partial a}{\partial \mu_Z} , \quad \frac{\partial a}{\partial \theta} = 2\rho a - \frac{1}{2}(\tilde{y}^2+\tilde{z}^2) . \tag{A.1}$$

Now suppose that F has an elliptical density of the form (4.2), and consider first the distribution of $a_{\theta,w}(X)$. Transform $(\tilde{y},\tilde{z})$ to

$$u,v = \frac{1}{2}\{b_1(\tilde{y}+\tilde{z}) \pm b_2(\tilde{y}-\tilde{z})\}$$

where $b_i = \{1-(-1)^i\rho\}^{-\frac{1}{2}}$. Then the joint density of (U,V) is

$$(1-\rho^2)^{\frac{1}{2}}\sigma_Y\sigma_Z k(|\Sigma|)\ \phi(u^2+v^2) , \tag{A.2}$$

and

$$a_{\theta,w}(X) = UV ,$$

which is the product of symmetric variables and hence symmetrically distributed.

Next we notice that

$$\frac{\partial a}{\partial \mu_Y} = \frac{1}{2\sigma_Y} \{(b_1+b_2) V + (b_1-b_2)U\}$$

so that the first component of $c_{\theta,w}$ in (3.4) is

$$E\{\frac{\partial a}{\partial \mu_Y} \dot{\psi}(a)\} = \text{constant} \times E\{V\dot{\psi}(UV)\} + \text{constant} \times E\{U\dot{\psi}(UV)\} ,$$

which is zero by the symmetry of (A.2) and the assumed symmetry of $\dot{\psi}$. The other components of $c_{\theta,w}$ are similarly found to be zero. Finally, the derivative $\partial a/\partial \theta$ in (A.1) is equal to

$$-\frac{1}{2}(U^2+V^2 - 2\rho UV) ,$$

so that $d_{\theta,w}$ in (3.5) can be written

$$d = (1-\rho^2)^{\frac{1}{2}} \sigma_Y \sigma_Z k(|\Sigma|) \iint u^2 \dot{\psi}(uv)\phi(u^2+v^2)dudv .$$

For any mixture of normal distribution $k(|\Sigma|) = k(1)\{\sigma_Y^2\sigma_Z^2(1-\rho^2)\}^{-\frac{1}{2}}$, in which case

$$d = k(1)\iint u^2\dot{\psi}(uv)\phi(u^2+v^2)dudv ,$$

independent of θ and w .